RM Misc. - Agree.
Thesis/
Reports
Brown,
B.G.

5120

Statistics and Climatic Impacts Laboratory (SCIL)

Department of Atmospheric Sciences

SCIL Report 85-2
(USDA-FS-RMFRES 28-K3-302)

VERIFICATION OF EXPERIMENTAL PROBABILISTIC SPOT FIRE-WEATHER FORECASTS

Barbara G. Brown

Allan H. Murphy

May 1985

Oregon State University

Corvallis, Oregon 97331

Statistics and Climatic Impacts Laboratory (SCIL) Reports

This series of reports will contain scientific papers, progress and project reports, and other special reports in the areas of statistical and applied meteorology and climatology prepared by members of the SCIL staff and/or by individuals associated with SCIL projects. These reports are intended to make the results of SCIL activities more widely available and easier to reference. Copies of the reports will be provided on a first-come, first-serve basis until supplies are exhausted.

Allan H. Murphy
Director, SCIL

Note

This report (SCIL 85-2) represents the fourth in a series of reports prepared under a project sponsored by the Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

VERIFICATION OF EXPERIMENTAL PROBABILISTIC SPOT FIRE-WEATHER FORECASTS

Report #4

to

Rocky Mountain Forest and Range Experiment Station
Forest Service, USDA
Fort Collins, Colorado 80523

by

Barbara G. Brown[1] and Allan H. Murphy

Department of Atmospheric Sciences
Oregon State University
Corvallis, Oregon 97331

May 1985

---

[1] Current address: Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, Colorado 80307.

Abstract

This report describes a study of the quality of - and uncertainty associated with - spot fire-weather forecasts prepared by National Weather Service (NWS) forecasters. The study involved the formulation of experimental probabilistic spot forecasts at four NWS offices for prescribed burns in a selected set of National Forests during the summer and fall of 1984. Experimental forecasts were prepared for four variables; namely, temperature, wind speed, wind direction, and relative humidity. Verification data were recorded by Forest Service personnel at the burn site at the planned ignition time (the valid time of the forecast).

The forecasters formulated a total of 178 experimental forecasts. However, matching observations were obtained on only 30% (53) of these occasions. This small set of probabilistic forecasts was evaluated using traditional measures of the quality of such forecasts (e.g., reliability diagrams, skill scores). In addition, the entire sample of 178 forecasts was analyzed to obtain some information regarding the overall characteristics of the forecasts.

The results of the experiment reveal that the probabilistic temperature forecasts were quite reliable and fairly precise. However, the lead times associated with these forecasts generally were very short and the forecasters usually had access to a very recent temperature observation from the burn site. These conditions may have contributed significantly to the forecasters' relatively successful efforts to formulate such forecasts. On the other hand, the experimental wind speed, wind direction, and relative humidity forecasts were considerably less successful. In particular, the skill scores for these forecasts

(based on sample climatology) were all negative, with especially large negative skill associated with the wind speed forecasts. Moreover, the forecasts were generally quite unreliable; specifically, they exhibited considerable overforecasting for most events.

Several factors may have contributed to the relatively poor quality of the experimental probabilistic spot fire-weather forecasts. These factors include: (1) a very small set of matched forecasts and observations; (2) the forecasters' lack of experience in quantifying the uncertainty in their spot forecasts; (3) the absence of any feedback to the forecasters regarding their performance during the experimental period; and (4) the possible lack of representativeness of some observations corresponding to the wind speed forecasts. Despite the somewhat disappointing results of this initial experiment in probabilistic spot fire-weather forecasting, it is proposed that further experiments of this type be undertaken, and several recommendations are made in an effort to enhance the likelihood of success in these future experiments.

Contents

Section Number and Title		Page
1.	Introduction	1
2.	Background and Objectives	2
3.	The Experiment	4
4.	Verification Methods	7
	a. Temperature forecasts	7
	b. Wind speed, wind direction, and relative humidity forecasts	8
5.	Results	11
	a. All forecasts	11
	b. Matched temperature forecasts	15
	c. Matched wind speed, wind direction, and relative humidity forecasts	30
6.	Summary and Discussion	44
7.	Conclusions and Recommendations	47
	Acknowledgments	50
	References	51
	Appendix	53

1. Introduction

This report is the fourth and final report in a series prepared under a project sponsored by the Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture. The purpose of the project was to develop procedures to quantify the uncertainty in spot fire-weather forecasts. This objective has been approached in two ways: (a) an investigation of the uncertainty in fire-danger forecasts [described in an earlier report; see Brown and Murphy (1985)] and (b) an experiment in which National Weather Service (NWS) forecasters explicitly quantified the uncertainty in spot forecasts for prescribed burns.

The purpose of this report is to describe the results of the spot fire-weather forecasting experiment. Some background information is presented in Section 2 and the experiment itself is described in Section 3. The methods used to verify the experimental forecasts are briefly outlined in Section 4 and the results of the experiment are summarized in Section 5. A discussion of the results and their implications is included in Section 6. Section 7 contains some conclusions based on this study of the probabilistic spot fire-weather forecasts and some recommendations regarding future work in this area.

2. Background and Objectives

Decisions regarding whether or not to proceed with a prescribed burn plan depend upon the assimilation of various types of information concerning factors that are of importance in determining the outcome of the proposed burn (i.e., factors influencing whether the burn meets its objectives, whether an unexpected loss of property or life occurs, the costs associated with the burn, etc.). Among the most important of these factors are the characteristics of the weather - including the humidity, temperature, and wind speed and direction - at the time that the burn is ignited and throughout the burn period. Information concerning these (and other) weather factors are obtained by the burn manager from specialized NWS fire-weather forecasters (or, in some cases, from private weather forecasters) at various times preceding and during the burn. These forecasts are termed "spot forecasts" because they represent forecasts for a particular location at a specific time.

Currently spot fire-weather forecasts are formulated and issued by the NWS forecasters as categorical forecasts [see Brown and Murphy (1984) for a description of the procedures and practices presently employed in formulating spot fire-weather forecasts]. That is, the burn managers are not provided with explicit information regarding the uncertainty in the forecasts. However, this information should be quite useful to individuals who must make decisions in which the uncertainty associated with the forecasts is of critical importance. Moreover, the issuance of probabilistic forecasts rather than categorical forecasts would allow the burn manager to make use of a formal procedure such as decision analysis in reaching important decisions related to whether or not to proceed with a planned prescribed burn (Radloff and Yancik, 1983).

The specific objective of this study is to evaluate the ability of fire-weather forecasters to formulate spot forecasts for prescribed burns in probabilistic terms. Spot forecasts include variables for which little experience exists in formulating probabilistic forecasts. Moreover, spot forecasts are issued for different remote locations on each occasion and these locations are typically in very complex or mountainous terrain. Hence, the forecasters generally have had no previous experience in forecasting for any of the burn locations, and they must make assumptions regarding the effects of the terrain on the values of the forecast variables. In combination, these factors produce a unique probability forecasting problem and contribute to making the formulation of probabilistic forecasts in this situation more difficult than it is in many other situations.

In order to assess the ability of fire-weather forecasters to quantify the uncertainty in spot forecasts, a probabilistic spot fire-weather forecasting experiment was undertaken during the summer and fall of 1984. In this experiment, fire-weather forecasters at four NWS offices prepared special probabilistic spot forecasts for four variables in conjunction with prescribed burns in a selected set of National Forests. The design of the experiment is described in the following section.

3. The Experiment

The experiment involved the formulation of probabilistic spot fire-weather forecasts for prescribed burns in a selected set of National Forests. Forecasters at the NWS offices in Boise (Idaho), Medford (Oregon), Missoula (Montana), and Wenatchee (Washington) participated in the experiment. The forecasts indicated the likelihood that certain temperature, wind speed, wind direction, and relative humidity events would occur at the planned ignition time of individual prescribed burns. These experimental forecasts were prepared in conjunction with the operational spot forecasts provided on request to the burn managers. Such forecasts are typically requested 1 or 2 hours prior to the planned ignition time and they are generally formulated and provided within 1/2 to 1 hour. Hence, most of the experimental forecasts had approximately 1/2 to 1 hour lead times.

Verifying observations were made at the burn site by Forest Service personnel, at approximately the time reported to the forecasters as the planned ignition time. These observations were presumably taken at the same location where an earlier observation - reported to the forecasters at the time that the forecast was requested - was made. Unfortunately, it was not possible to confirm whether or not this condition was always satisfied (however, we believe that it did hold in general). The observations were mailed directly to the authors by the observers or their associates.

As indicated above, experimental probabilistic forecasts were formulated for temperature (T), wind speed (WS), wind direction (WD), and relative humidity (RH). In the case of temperature, each forecast involved the assessment of the median (T50), lower quartile (T25), and upper quartile (T75) of the forecaster's subjective probability distribution.

The difference between T75 and T25 represents a 50% central credible interval temperature forecast (these intervals are assumed to include their endpoints). In the case of wind speed and relative humidity, probabilities were assessed for three overlapping events. Specifically, probability forecasts were formulated for WS > 5 mph (WGT5), WS > 10 mph (WGT10), and WS > 15 mph (WGT15), and for RH < 40% (HLT40), RH < 30% (HLT30), and RH < 20% (HLT20). The threshold values for these forecasts were selected after extensive discussions with forecasters at the four offices. For the wind direction forecasts, the forecasters first formulated their best estimate of the wind direction (WD) to the nearest of 8 cardinal compass points and then assessed their subjective probability that the observed wind direction would be within ±45° of this value (PWD). Permissible probability values used in the wind speed, wind direction, and relative humidity forecasts included the fifteen values 0.00, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.98, and 1.00. Temperatures were assessed to the nearest degree Fahrenheit.

The individuals who participated in the experiment were experienced weather forecasters, and they were regular participants in the operational spot fire-weather forecasting program. However, the forecasters had relatively little experience in quantifying the uncertainty in spot forecasts. In fact, their probability forecasting experience consisted exclusively of the preparation of a relatively small number of probability of precipitation forecasts in conjunction with the NWS public weather forecasting program. In an effort to overcome this problem, some materials were prepared to aid the forecasters in assessing the relevant probabilities and in checking the consistency of their forecasts (see Appendix), and they were encouraged to make some practice forecasts prior to the start of the

experiment. No formal feedback was provided to the forecasters during the experimental period.

The experiment was initiated on 16 April 1984 and it continued into the fall (the last forecast included in the study was formulated on 23 October 1984). A total of 178 forecasts was collected during this period, including 23 from Boise, 17 from Medford, 106 from Missoula, and 32 from Wenatchee. However, due to difficulties in obtaining verifying observations made at the appropriate times, the numbers of matching forecasts and observations are much smaller than these total values. After eliminating forecasts for which no verifying observation existed and those for which the observation was not made at a time near the expected verification time, a total of only 53 forecast-observation pairs remains (37 pairs from Missoula and 16 pairs from the other offices). For many events, the numbers of cases are even less than these values due to missing observations or to the accidental omission of the forecast by the forecaster.

4. Verification Methods

Several different methods and measures are used to evaluate the statistical characteristics and performance of the forecasts formulated in the spot fire-weather forecasting experiment. Since the temperature forecasts differ markedly in form and format from the wind speed, wind direction, and relative humidity forecasts, we discuss the verification methods for these two types of forecasts separately. For a more detailed treatment of these and other methods of evaluating probability forecasts, refer to Murphy and Daan (1984, 1985) and Winkler and Murphy (1979).

a. Temperature forecasts

It is of some interest to examine the distributions of the temperature forecasts actually assessed by the forecasters (i.e., T25, T50, and T75) and to compare the distribution of T50 with the distribution of observed temperatures (T). For this purpose, statistics such as the mean, median, and standard error (of the mean) of the respective distributions can be calculated.

The performance of the T50 forecasts as point forecasts can be evaluated in several different ways. For example, a scatter diagram of T50 versus T can be plotted, and the points in such a diagram can be compared with the 45° line T50 = T. The overall bias of the T50 forecasts can be assessed by determining the relative frequency of observed temperatures less than, equal to, and greater than these median temperatures and by computing the mean error (ME) of these forecasts. Suitable measures of the accuracy of the T50 forecasts include the mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (CC).

With regard to the reliability of the 50% central credible intervals (i.e., the interval from T25 to T75, inclusive), a reasonable approach is to

determine the relative frequencies of observed temperatures below, in, and above the intervals and to compare these relative frequencies with the "theoretical" values (i.e., 0.25, 0.50, 0.25). When such relative frequencies correspond closely to the respective theoretical probabilities, the forecasts are considered to be reliable; otherwise, the forecasts are considered to be unreliable to some extent. Another important attribute of credible interval temperature forecasts is their precision, and a natural measure of this attribute of the forecasts is their average width.

The asymmetry of the 50% central credible interval forecasts is of some interest, and several measures of asymmetry can be defined. For example, the mean asymmetry (MA) can be determined as the average difference between T50-T25 and T75-T50. Another measure of asymmetry is the mean absolute asymmetry (MAA), where MAA is the average absolute difference between T50-T25 and T75-T50.

It is possible to obtain a composite measure of the reliability and precision of variable-width credible interval forecasts by computing the average "losses", where these losses are considered to represent the average expenses incurred by an individual who makes decisions on the basis of the forecasts. A loss function of this type has been described by Winkler and Murphy (1979), and it is used here to evaluate the 50% central credible interval forecasts. In this case, the loss is equal to the width of the interval when the observed temperature falls in the interval and it equals the sum of this width and a "distance" penalty when the observed temperature falls below or above the interval.

b. Wind speed, wind direction, and relative humidity forecasts

In the case of the wind speed, wind direction, and relative humidity forecasts, overall reliability (or bias) can be determined by comparing the

average forecast probabilities and observed relative frequencies for the respective events. The difference between such probabilities and relative frequencies is a natural measure of this attribute of the forecasts.

Reliability diagrams provide a means of assessing the extent to which the experimental probability forecasts possess this important and desirable attribute. In these diagrams, observed relative frequency is plotted against forecast probability for specific probability values (i.e., the permissible values). The sample curve joining these points is then compared with the diagonal 45° line which represents complete correspondence between forecast probabilities and observed relative frequencies. In judging the significance of departures from this perfect reliability line, it is important to keep in mind that the sampling variability associated with a particular point on the empirical curve is roughly inversely proportional to the square root of the size of the respective sample of forecasts.

A natural measure of the accuracy of probability forecasts for ordinal variables such as wind speed and relative humidity is the ranked probability score (RPS) (Epstein, 1969; Murphy, 1971). In effect, this measure is simply the mean square error of (cumulative) probabilistic forecasts (thus, smaller scores are better). When only a finite number of probability values are used, it is possible to partition the RPS into three terms: (1) a term representing the RPS for forecasts based solely on the sample climatological probabilities (RPS_{SC}), (2) a measure of reliability (REL), and (3) a measure of resolution (RES). Specifically, $RPS = RPS_{SC} + REL - RES$. To determine the skill of a sample of forecasts, it is traditional to compare the accuracy of these forecasts with the accuracy of forecasts based on some standard of reference such as climatology. Thus, a natural skill score (SS) in this context is $SS = 100[1 - (RPS/RPS_{SC})] = 100[(RES - REL)/RPS_{SC}]$. SS represents the

improvement (in percent) in the accuracy of the forecasts of interest over the accuracy of forecasts based on sample climatology.

The Brier score (BS) is used as a measure of the accuracy of the wind direction probability forecasts. (It should be noted that the Brier score is equivalent to the ranked probability score in the two-event situation.) This score can be partitioned - in a manner similar to the RPS - into BS_{SC} (a term representing the BS based solely on the sample climatological probabilities), REL, and RES, and a skill score (SS) based on sample climatology can be computed.

5. Results

In this section we describe the results of the experiment. Because the number of matched forecasts and observations is relatively small, it would appear to be of some value to examine the characteristics of all of the experimental forecasts (whether or not they had verifying observations), as well as to evaluate the matched forecasts. Hence, Section 5a consists of a brief description of the characteristics of all 178 forecasts that were formulated during the experiment, and Sections 5b and 5c contain the verification results for the 53 matched forecasts and observations.

a. All forecasts

A summary of the characteristics of the experimental temperature forecasts is presented in Table 1, including the means, medians, and standard errors of the forecast temperature quantiles. The forecasts are stratified by NWS office and, for Missoula, by forecaster (note that Missoula was the only office with a sufficient number of forecasts to allow stratification by forecaster). Table 1 indicates that, on the average, the Medford temperature forecasts were the lowest and the Missoula forecasts were the highest. Otherwise, the differences among the offices and forecasters are not notable.

The average widths of the forecast temperature intervals are summarized in Table 2. These values (particularly the width of the central interval from T25 to T75) provide an indication of the precision of the forecasts. In particular, it appears that the intervals were narrowest at Missoula and widest at Medford. Among the Missoula forecasters, Forecaster 1 (Forecaster 3) provided the narrowest (widest) intervals.

Table 3 contains a summary of the temperature asymmetry measures MA and MAA. It appears from the mean asymmetry (MA) values that the forecasts

Table 1. Statistics of temperature quantiles (all in °F) for all forecasts.

Office	Sample size	Statistic	Temperature quantile T25	T50	T75
All	178	Mean	63.2	66.3	68.9
		Median	63	67	69
		Standard error	0.79	0.79	0.78
Boise	23	Mean	62.2	65.7	68.6
		Median	63	64	66
		Standard error	2.60	2.68	2.68
Medford	17	Mean	59.1	62.5	65.7
		Median	59	61	63
		Standard error	2.47	2.56	2.50
Missoula	106	Mean	65.1	67.9	70.4
		Median	65	67	70
		Standard error	0.94	0.94	0.92
Wenatchee	32	Mean	59.8	63.2	66.0
		Median	58	62.5	66.5
		Standard error	1.91	1.93	1.94
Missoula: Forecaster 1	24	Mean	65.0	67.4	69.8
		Median	64.5	67	69.5
		Standard error	1.97	2.06	2.04
Missoula: Forecaster 2	40	Mean	64.8	67.4	69.9
		Median	65.5	68	70
		Standard error	1.60	1.62	1.58
Missoula: Forecaster 3	15	Mean	61.6	65.0	67.4
		Median	60	65	67
		Standard error	2.35	2.11	2.11
Missoula: Forecaster 4	25	Mean	69.0	72.1	74.4
		Median	68	73	74
		Standard error	1.62	1.58	1.57

Table 2. Average widths of forecast temperature intervals (standard errors in parentheses) for all forecasts. All values in °F.

Office	Sample size	Average widths T25-to-T50	T50-to-T75	T25-to-T75
All	178	4.0(0.12)	3.7(0.09)	6.7(0.14)
Boise	23	4.4(0.41)	3.9(0.34)	7.3(0.65)
Medford	17	4.4(0.32)	4.2(0.25)	7.6(0.38)
Missoula	106	3.8(0.14)	3.5(0.10)	6.3(0.14)
Wenatchee	32	4.4(0.27)	3.8(0.18)	7.3(0.33)
Missoula: Forecaster 1	24	3.4(0.26)	3.5(0.21)	5.9(0.26)
Missoula: Forecaster 2	40	3.6(0.22)	3.5(0.13)	6.1(0.24)
Missoula: Forecaster 3	15	4.4(0.43)	3.4(0.27)	6.8(0.47)
Missoula: Forecaster 4	25	4.1(0.31)	3.3(0.22)	6.4(0.24)

Table 3. Measures of asymmetry of 50% central credible interval temperature forecasts for all forecasts. MA = mean asymmetry, MAA = mean absolute asymmetry. Standard errors appear in parentheses.

Office	Sample size	MA (°F)	MAA (°F)
All	178	0.38(0.15)	1.43(0.10)
Boise	23	0.52(0.39)	1.48(0.26)
Medford	17	0.12(0.43)	1.29(0.28)
Missoula	106	0.31(0.20)	1.52(0.14)
Wenatchee	32	0.63(0.31)	1.19(0.25)
Missoula: Forecaster 1	24	-0.04(0.40)	1.46(0.26)
Missoula: Forecaster 2	40	0.10(0.27)	1.00(0.22)
Missoula: Forecaster 3	15	1.00(0.55)	1.93(0.33)
Missoula: Forecaster 4	25	0.80(0.49)	2.08(0.29)

provided by the Medford forecasters and by Forecaster 1 at Missoula were symmetric on the average. However, the values of MAA indicate that the lowest mean absolute values of asymmetry are associated with the Wenatchee forecasts and with Forecaster 2 at Missoula.

Some statistics for the wind speed forecasts are summarized in Table 4. The means and medians of the forecast probabilities decrease monotonically from WGT5 to WGT10 to WGT15 for all offices and forecasters. In addition, the mean probabilities for WGT15 are quite small (ranging from approximately 0.036 to 0.118), reflecting the relative rarity of this event. On the other hand, the average probabilities for WGT5 are quite large.

The relative frequencies of use of the eight wind directions are presented in Table 5. Overall, the most frequently forecast direction is southwest, and this result holds for each of the individual offices as well. Table 6 contains a summary of the probability values used in the wind direction forecasts, including the means, medians, and standard errors of these forecasts. The most notable feature of this table is the size of the average probabilities. In all cases except the mean probability for Forecaster 3 at Missoula, the mean and median values of PWD are 0.60 or greater.

Statistics for the humidity forecasts are summarized in Table 7. The mean and median values of these probabilities increase monotonically from HLT20 to HLT30 to HLT40 for all offices and forecasters. In general, the standard errors of the probabilities also increase in a similar manner.

b. Matched temperature forecasts

Although primary interest is focused on the temperature intervals as indicators of the uncertainty in the spot forecasts, we first consider the median temperatures as point forecasts. Scatter diagrams of median

Table 4. Statistics of wind speed probability forecasts for all forecasts.

Office	Sample size	Statistic	Wind speed events WGT5	WGT10	WGT15
All	178	Mean	0.532	0.203	0.075
		Median	0.500	0.200	0.050
		Standard error	0.016	0.014	0.010
Boise	23	Mean	0.504	0.250	0.078
		Median	0.500	0.200	0.050
		Standard error	0.040	0.036	0.019
Medford	17	Mean	0.547	0.232	0.082
		Median	0.600	0.200	0.050
		Standard error	0.070	0.048	0.024
Missoula	106	Mean	0.550	0.206	0.085
		Median	0.500	0.200	0.050
		Standard error	0.020	0.019	0.015
Wenatchee	32	Mean	0.484	0.147	0.036
		Median	0.500	0.100	0.020
		Standard error	0.031	0.019	0.008
Missoula: Forecaster 1	24	Mean	0.554	0.151	0.036
		Median	0.550	0.100	0.020
		Standard error	0.038	0.022	0.012
Missoula: Forecaster 2	40	Mean	0.514	0.221	0.118
		Median	0.500	0.200	0.050
		Standard error	0.033	0.036	0.034
Missoula: Forecaster 3	15	Mean	0.450	0.173	0.075
		Median	0.500	0.200	0.050
		Standard error	0.042	0.030	0.012
Missoula: Forecaster 4	25	Mean	0.640	0.238	0.077
		Median	0.700	0.200	0.000
		Standard error	0.037	0.046	0.027

Table 5. Relative frequencies of use of eight wind directions for all forecasts.

Office	Sample size	Wind direction N	NE	E	SE	S	SW	W	NW
All	177	0.090	0.068	0.079	0.096	0.136	0.249	0.175	0.107
Boise	23	0.130	0.043	0.043	0.130	0.174	0.217	0.174	0.087
Medford	17	0.235				0.118	0.412	0.176	0.059
Missoula	105	0.076	0.076	0.095	0.105	0.124	0.238	0.171	0.114
Wenatchee	32	0.031	0.094	0.094	0.094	0.156	0.219	0.188	0.125
Missoula: Forecaster 1	23		0.087	0.087	0.261	0.130	0.261	0.174	
Missoula: Forecaster 2	40	0.050	0.100	0.025	0.075	0.050	0.225	0.275	0.200
Missoula: Forecaster 3	15	0.067		0.200	0.067	0.333	0.133	0.067	0.138
Missoula: Forecaster 4	25	0.200		0.160	0.040	0.120	0.320	0.080	0.080

Table 6. Statistics of wind direction probability forecasts for all forecasts.

Office	Sample size	Mean	Median	Standard error
All	169	0.670	0.700	0.011
Boise	23	0.722	0.800	0.030
Medford	17	0.647	0.700	0.054
Missoula	97	0.654	0.700	0.014
Wenatchee	32	0.697	0.700	0.021
Missoula: Forecaster 1	23	0.626	0.600	0.016
Missoula: Forecaster 2	40	0.718	0.750	0.022
Missoula: Forecaster 3	15	0.547	0.600	0.024
Missoula: Forecaster 4	19	0.637	0.700	0.033

Table 7. Statistics of relative humidity probability forecasts for all forecasts.

Office	Sample size	Statistic	Relative humidity events HLT20	HLT30	HLT40
All	178	Mean	0.056	0.217	0.506
		Median	0.000	0.100	0.500
		Standard error	0.009	0.020	0.025
Boise	23	Mean	0.154	0.376	0.617
		Median	0.100	0.400	0.600
		Standard error	0.037	0.062	0.049
Medford	17	Mean	0.050	0.106	0.321
		Median	0.000	0.000	0.200
		Standard error	0.047	0.060	0.078
Missoula	106	Mean	0.034	0.195	0.482
		Median	0.000	0.100	0.500
		Standard error	0.009	0.026	0.032
Wenatchee	32	Mean	0.059	0.236	0.607
		Median	0.020	0.200	0.600
		Standard error	0.018	0.040	0.058
Missoula: Forecaster 1	24	Mean	0.013	0.163	0.398
		Median	0.000	0.100	0.300
		Standard error	0.004	0.037	0.065
Missoula: Forecaster 2	40	Mean	0.045	0.166	0.394
		Median	0.010	0.075	0.300
		Standard error	0.017	0.041	0.052
Missoula: Forecaster 3	15	Mean	0.020	0.187	0.607
		Median	0.000	0.100	0.600
		Standard error	0.014	0.059	0.060
Missoula: Forecaster 4	25	Mean	0.044	0.248	0.614
		Median	0.000	0.000	0.700
		Standard error	0.022	0.069	0.071

temperature (T50) versus observed temperature (T) for all stations combined and for the Missoula forecasters alone are presented in Figures 1 and 2, respectively. These diagrams reveal that most of the points are quite close to the T50=T line, with only a few points exhibiting large departures from this line. That is, a close correspondence appears to exist between T50 and T.

Quantitative information concerning the bias (overforecasting/underforecasting) in the forecasters' median temperatures is contained in Table 8. These results indicate that the medians are relatively unbiased. For all offices, the observed temperature is greater than the median in 50% of the cases and less than the median in 42% of the cases.

Some measures of the performance of the median temperature forecasts for all offices combined and for Missoula are presented in Table 9. These verification statistics provide quantitative support for some of the statements made in the previous paragraphs regarding the quality of the T50 forecasts. For example, the values of the correlation coefficient are indicative of a relatively high degree of correspondence between T50 and T. Moreover, it can be seen that the forecasters' medians exhibit only a very slight bias toward underforecasting (ME=-0.4°F for all stations; ME=-0.5°F for Missoula). The MAE and RMSE values, ranging between 4 and 6°F, are quite reasonable for such forecasts.

Summary statistics for the observed temperature (T) and temperature quantile forecasts (T25, T50, and T75) are contained in Table 10. The Missoula means and medians are slightly larger than those for all offices combined. However, the standard errors of the temperature quantiles are equal for both sets of forecasts. It should also be noted that the standard errors

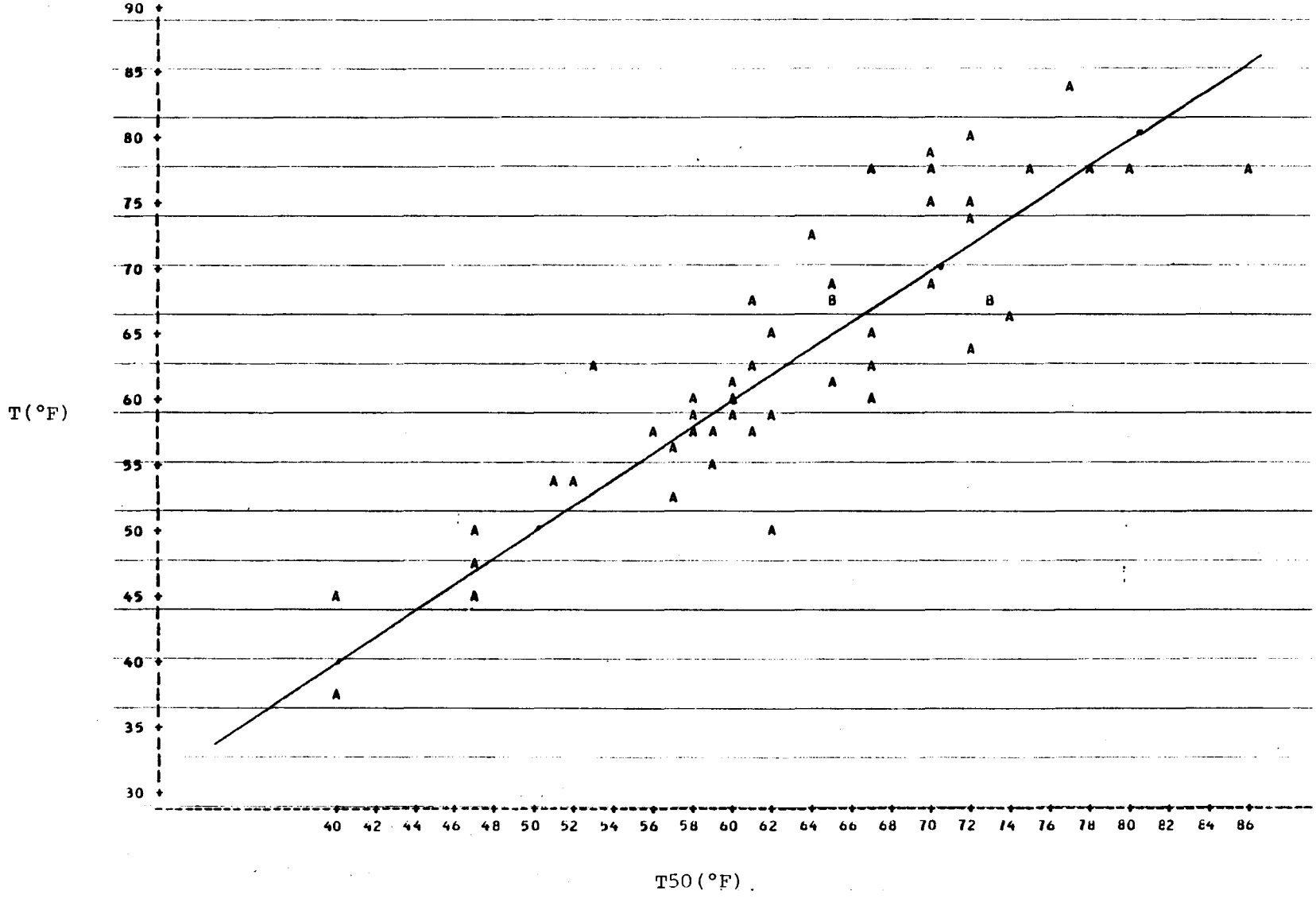

Figure 1. Scatter diagram relating T50 and T for all offices combined. A = 1 case, B = 2 cases.

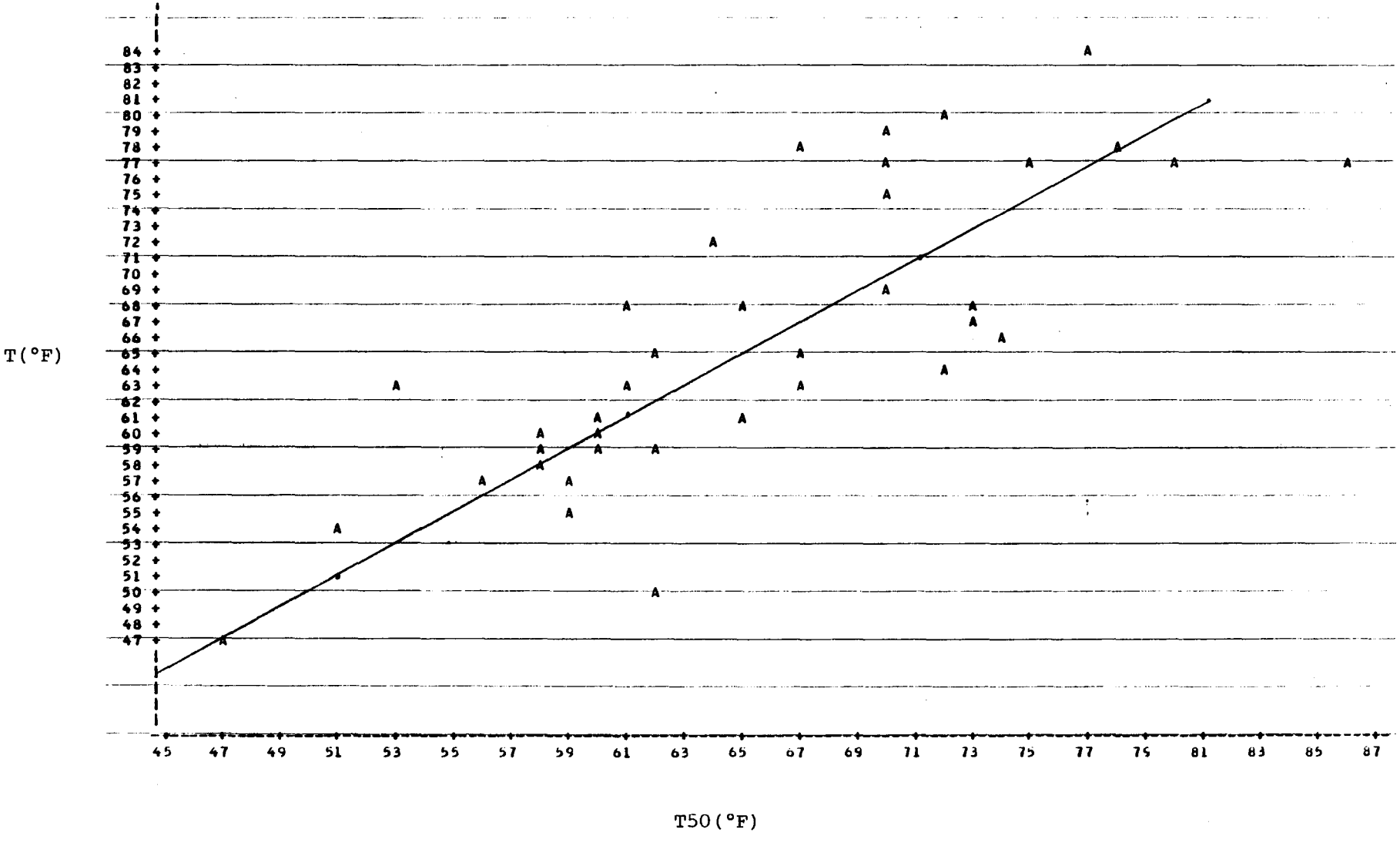

Figure 2. Scatter diagram relating T50 and T for Missoula forecasts. A = 1 case.

Table 8. Relative frequencies of observed temperatures less than, equal to, and greater than forecasters' median temperatures.

Office	Sample size	Relative frequencies T<T50	T=T50	T>T50
All	50	0.420	0.080	0.500
Missoula	37	0.405	0.108	0.486

Table 9. Measures of performance of forecasters' median temperatures as point forecasts. ME = mean error, MAE = mean absolute error, RMSE = root mean square error, CC = correlation coefficient.

Office	Sample size	ME (°F)	MAE (°F)	RMSE (°F)	CC
All	50	-0.4	4.1	5.1	0.87
Missoula	37	-0.5	4.4	5.5	0.80

Table 10. Statistics of observed temperature and temperature quantiles (all in °F) for matched forecasts.

Office	Sample size	Statistic	Observed temperature T	Temperature quantiles T25	T50	T75
All	50	Mean	63.7	60.6	63.3	66.1
		Median	63.0	60.0	63.0	66.0
		Standard error	1.5	1.4	1.4	1.4
Missoula	37	Mean	65.9	63.0	65.5	68.1
		Median	65.0	62.0	65.0	67.0
		Standard error	1.5	1.4	1.4	1.4

of the medians (T50) are less than the standard errors of the observed temperatures.

The relative frequencies of temperatures below, in, and above the interval from T25 to T75 are indicated in Table 11. These results characterize the reliability of the 50% central credible interval forecasts. The relative frequency in the T25 to T75 interval slightly exceeds the expected value of 0.500 for all offices combined and is somewhat less than 0.500 for the Missoula forecasts. This result suggests that, on the average, the intervals assessed by all of the forecasters combined were approximately the correct width, whereas the intervals assessed by the Missoula forecasters were somewhat too narrow. Apparently the forecasters at Boise, Medford, and Wenatchee used intervals that were somewhat too wide, on the average, which led to the overall result. The statistics concerning the width of the temperature intervals in Table 11 represent indicators of the precision of the forecasts. Overall, the 50% central credible intervals (T25 to T75) are approximately 6°F wide, and they do not exhibit much variability.

Measures of the asymmetry in the forecasters' 50% central credible intervals are reported in Table 12. These measures represent the average difference and average absolute difference between T50-T25 and T75-T50. Small values of such measures are indicative of symmetry, whereas large values of the measures are indicative of asymmetry. Overall, the T75-T50 intervals are larger than the T50-T25 intervals, as indicated by the slightly negative MA values. The intervals assessed by the Missoula forecasters are somewhat more asymmetric than those assessed by the forecasters at all four offices combined.

Statistics (mean, median, and standard error) of the loss function are presented in Table 13. The loss function is a composite measure of the relia-

Table 11. Relative frequencies of observed temperatures below, in, and above forecasters' 50% central credible interval temperature forecasts and mean and standard error of width of forecasters' intervals.

Office	Sample size	Relative frequencies			Mean width (standard error) (°F)
		T<T25	T25≤T≤T75	T>T75	
All	50	0.260	0.520	0.220	6.4(0.2)
Missoula	37	0.297	0.432	0.270	6.2(0.2)

Table 12. Measures of asymmetry of 50% central credible interval temperature forecasts for matched forecasts. MA = mean asymmetry, MAA = mean absolute asymmetry. Standard errors appear in parentheses.

Office	Sample size	MA (°F)	MAA (°F)
All	50	-0.16(0.21)	1.08(0.15)
Missoula	37	-0.19(0.27)	1.22(0.18)

Table 13. Statistics of loss function for forecasters' 50% central credible interval temperature forecasts.

Office	Sample size	Mean	Median	Standard error
All	50	14.1	9.0	1.5
Missoula	37	15.7	9.0	2.0

bility and precision of the credible interval forecasts (see Section 4a). These statistics reveal that the average loss associated with the Missoula forecast intervals exceeds the average loss associated with the intervals for the four offices combined. This result is related to the fact that the Missoula forecasters, on the average, tended to underforecast the width of the central credible intervals (see Table 11).

c. Matched wind speed, wind direction, and relative humidity forecasts

The average forecast probability, observed relative frequency, and overall reliability (or bias) of the probabilistic wind speed and wind direction forecasts are indicated in Table 14. Table 15 contains the values of these statistics for the probabilistic relative humidity forecasts. The statistics are presented for each wind speed, wind direction, and relative humidity event for all stations offices and for the Missoula forecasters alone.

With regard to overall reliability, the results for all offices combined and for Missoula alone indicate that observed relative frequency remains constant or increases as average forecast probability increases for all wind speed and relative humidity events. However, overforecasting exists for all of these events. In this regard, the events HLT20, WGT10, and WGT15 never occurred in the experimental period. The wind direction (PWD) forecasts for all offices combined exhibit a small amount of underforecasting, whereas the PWD forecasts for Missoula are almost completely reliable in-the-large.

Reliability diagrams for the probabilistic wind speed forecasts of the event WGT5 for all offices combined and for Missoula alone are presented in Figure 3. The inset in this figure contains the frequency-of-use distributions for the various forecast probability values. Note that results are presented only for those probability values that were used more than three

Table 14. Average forecast probability, observed relative frequency, and overall reliability for experimental wind speed and direction forecasts for (a) all offices combined and (b) Missoula alone.

Event	Sample size	Average forecast probability (standard error)	Observed relative frequency	Overall reliability
(a) All offices				
WGT5	46	0.537(0.031)	0.283	0.254
WGT10	46	0.189(0.026)	0.000	0.189
WGT15	46	0.067(0.021)	0.000	0.067
PWD	46	0.680(0.028)	0.739	-0.059
(b) Missoula				
WGT5	34	0.521(0.028)	0.353	0.168
WGT10	34	0.185(0.027)	0.000	0.185
WGT15	34	0.073(0.023)	0.000	0.073
PWD	30	0.673(0.025)	0.667	0.006

Table 15. Average forecast probability, observed relative frequency, and overall reliability for experimental relative humidity forecasts for (a) all offices combined and (b) Missoula alone.

Event	Sample size	Average forecast probability (standard error)	Observed relative frequency	Overall reliability
(a) All offices				
HLT20	53	0.034(0.015)	0.000	0.034
HLT30	53	0.165(0.034)	0.094	0.071
HLT40	53	0.432(0.045)	0.358	0.074
(b) Missoula				
HLT20	37	0.022(0.009)	0.000	0.022
HLT30	37	0.170(0.041)	0.108	0.062
HLT40	37	0.443(0.054)	0.351	0.092

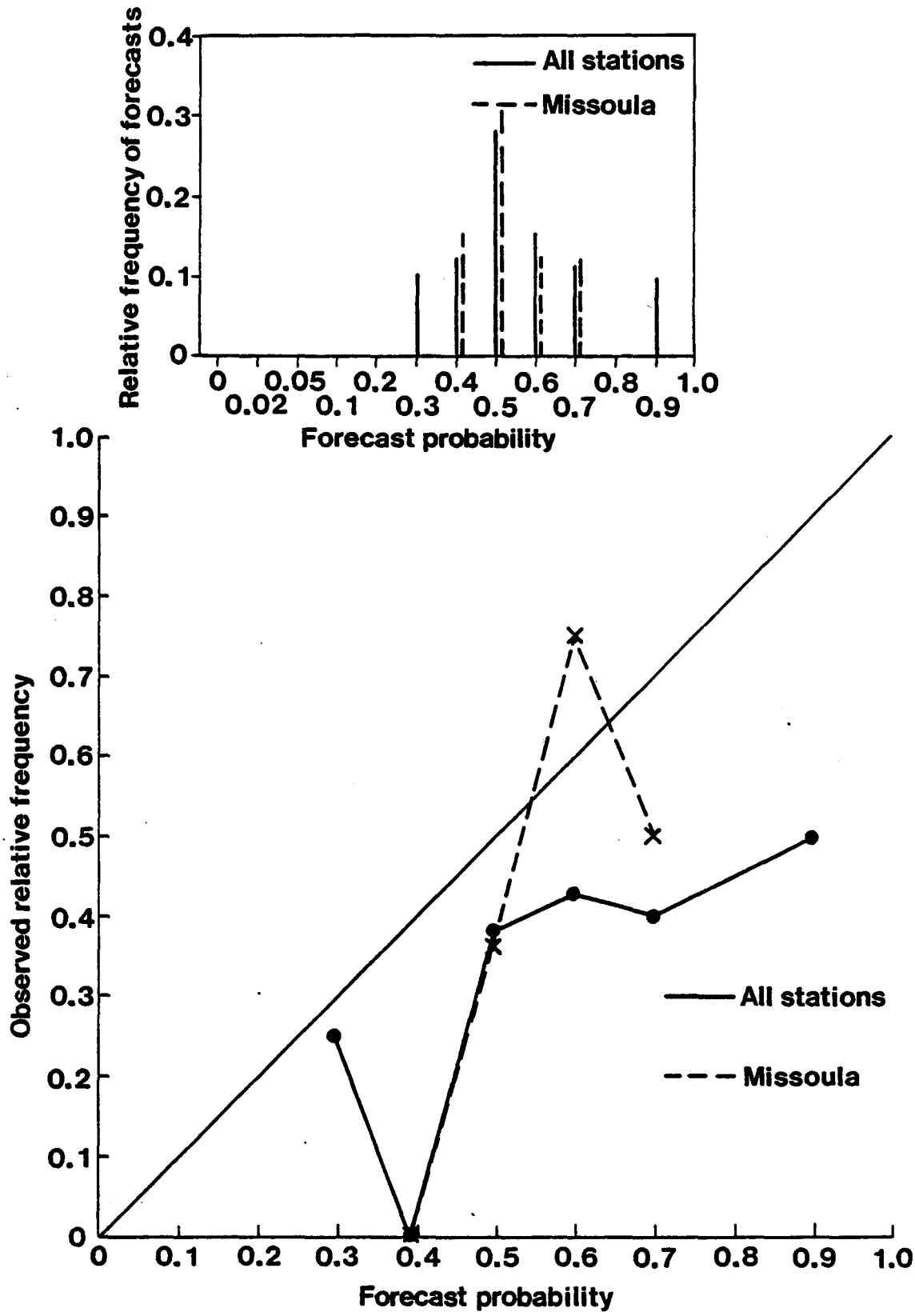

Figure 3. Reliability diagram for WGT5 forecasts. Inset represents frequency-of-use diagram.

times. Figure 3 reveals that the WGT5 forecasts exhibit considerable overforecasting. For WGT10 and WGT15, events that never occurred in the experimental period, it is of interest to examine the frequency-of-use diagrams shown in Figure 4. The diagram for WGT10 reveals that considerable overforecasting occurred for this event. In this regard, the mode of the distribution corresponds to a forecast probability of 0.10 and a probability value of 0.40 was used at least four times. On the other hand, the WGT15 forecasts (Figure 4b) are fairly reliable - the mode of the distribution for all stations combined is 0.00 and the largest forecast probability value that was used is 0.10.

The wind direction (PWD) reliability diagram is presented in Figure 5. The frequency-of-use distribution for PWD reveals that the forecast probability values for this event fall in a range between 0.50 and 0.90. Forecasts of 0.50, 0.60, and 0.80 are fairly reliable, but the relationship between observed relative frequency and forecast probability is not monotonic.

In the case of the relative humidity forecasts, the reliability diagrams for HLT40 and HLT30 appear in Figure 6. The HLT40 forecasts are quite reliable, with only a slight indication of underforecasting. However, the HLT30 forecasts are reliable only for low forecast probabilities, with the forecast probability value of 0.40 being associated with considerable overforecasting. The frequency-of-use diagram for HLT20 (another event that never occurred in the sample) presented in Figure 7 indicates that the HLT20 forecasts are actually very reliable, with approximately 70% of the forecast probabilities equal to 0.00. As in the case of the wind speed forecasts, the frequency-of-use distributions for the relative humidity forecasts tend to reflect the differences in sample relative frequencies of the events and the range of

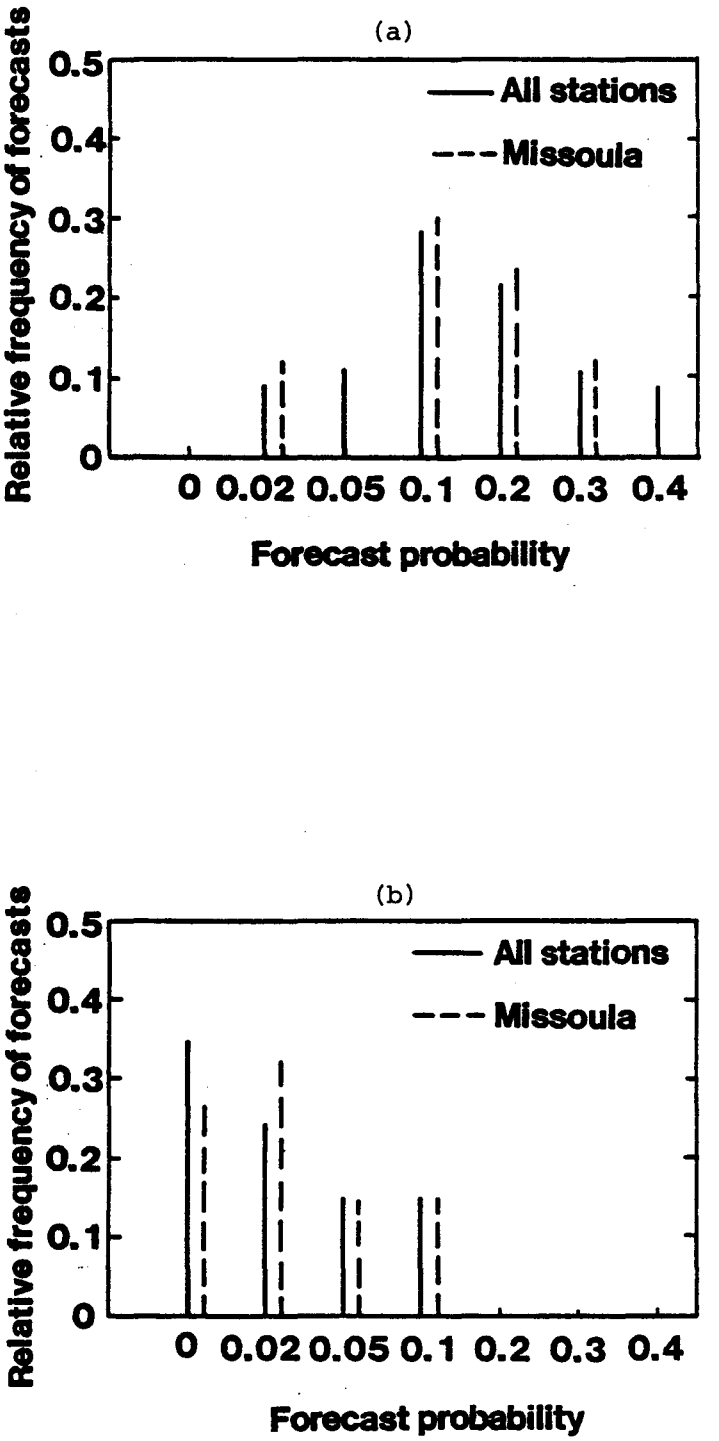

Figure 4. Frequency-of-use diagrams for (a) WGT10 forecasts and (b) WGT15 forecasts.

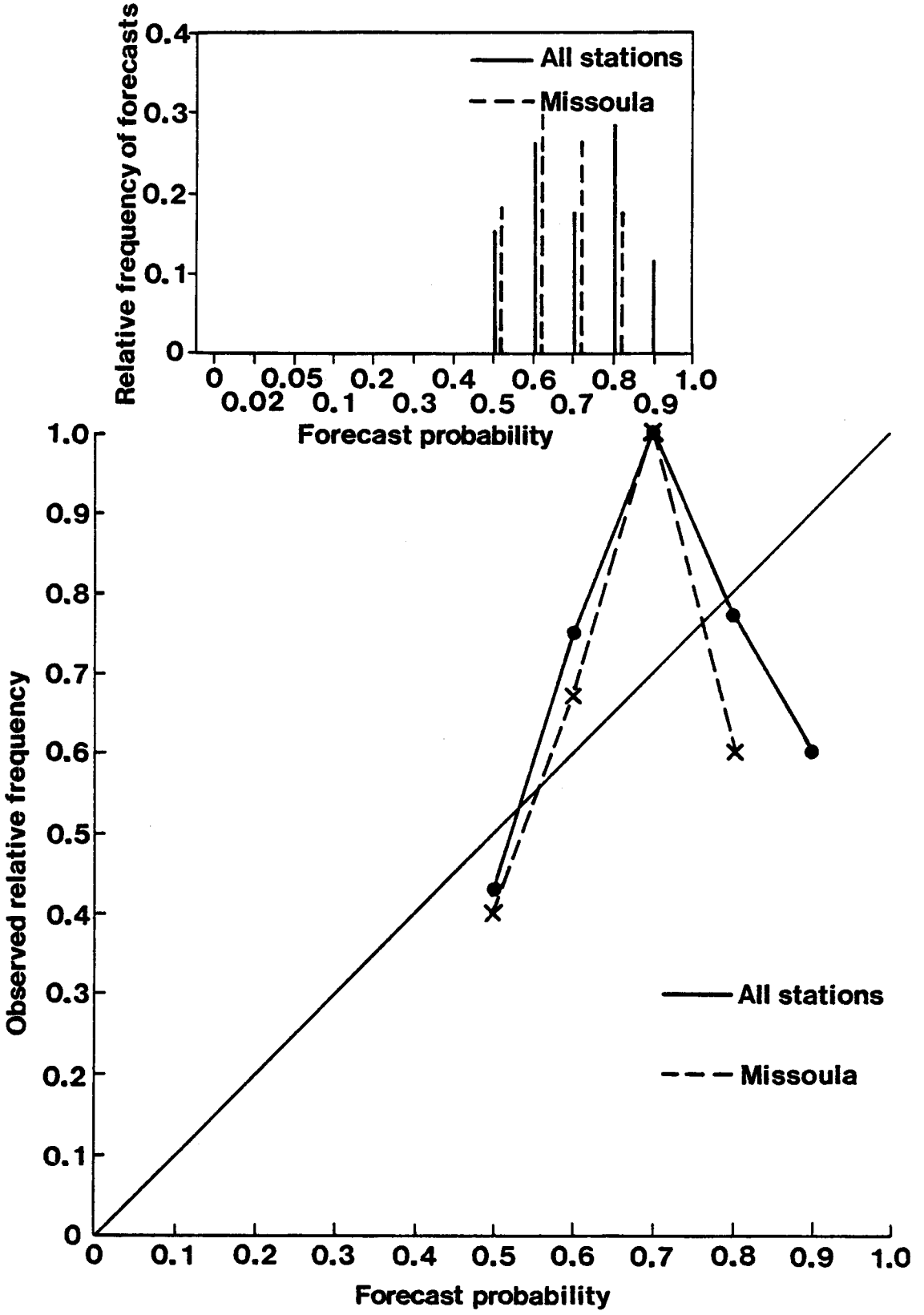

Figure 5. Reliability diagram for PWD forecasts. Inset represents frequency-of-use diagram.

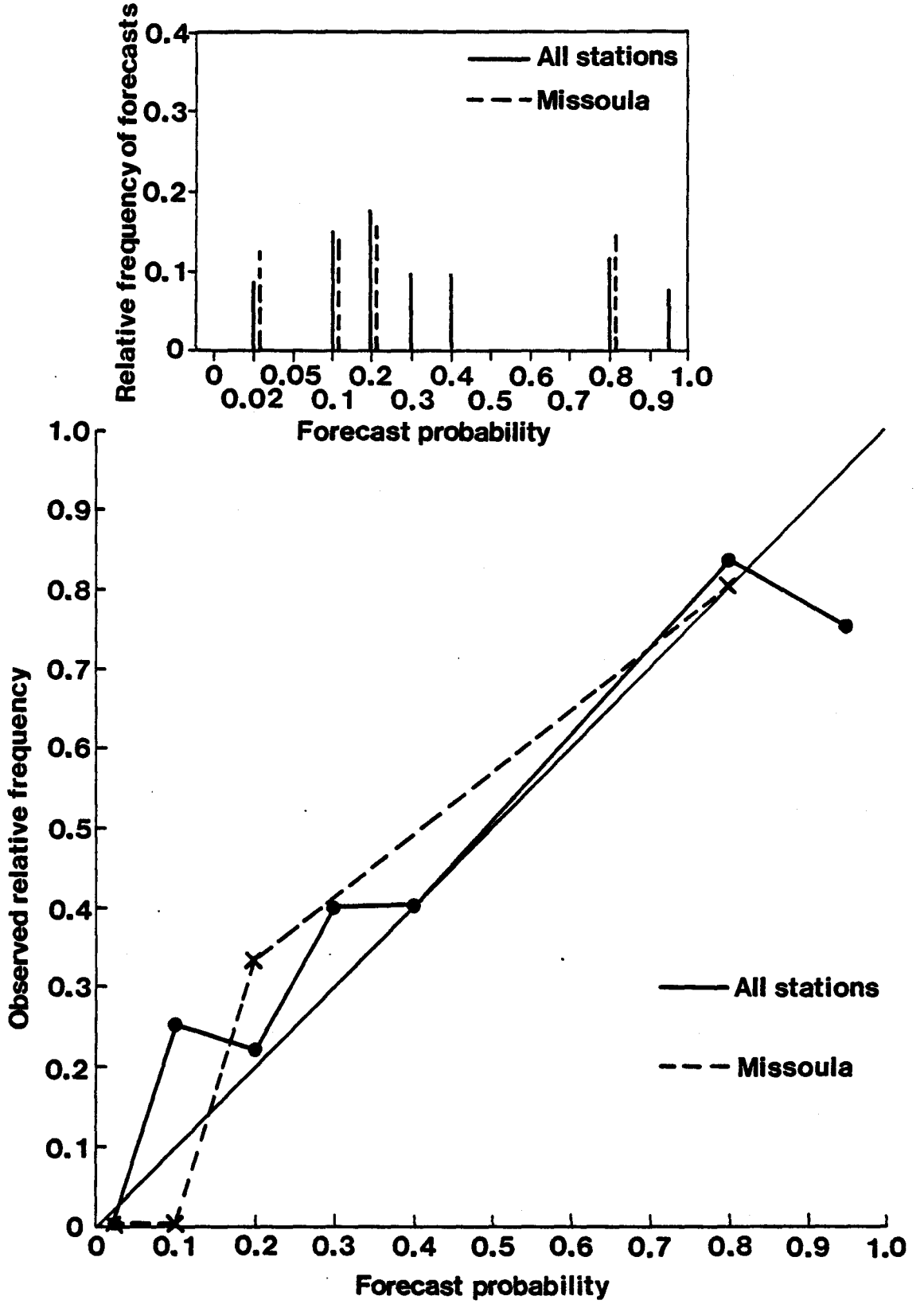

Figure 6a. Reliability diagram for HLT40 forecasts. Inset represents frequency-of-use diagram.

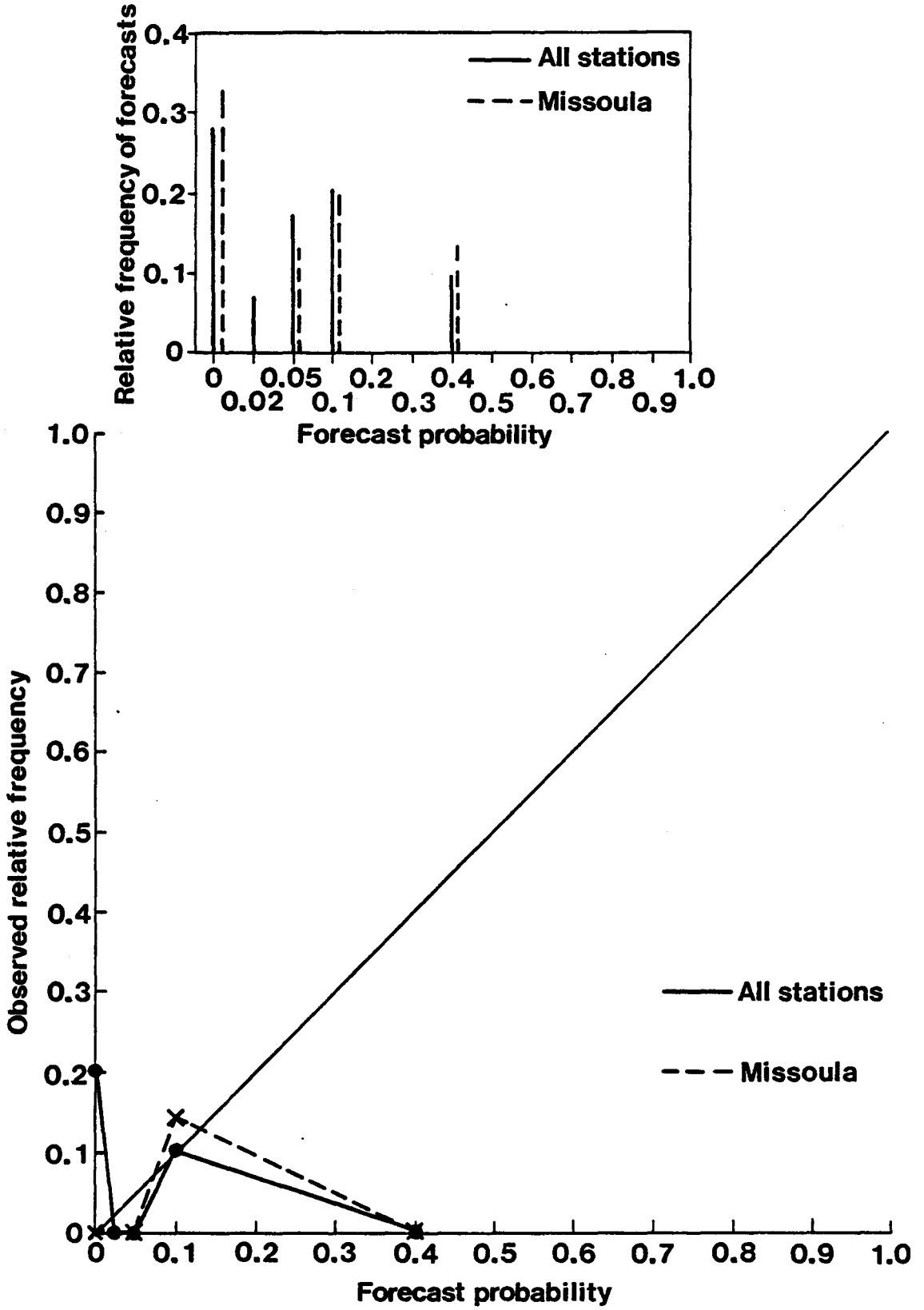

Figure 6b. Reliability diagram for HLT30 forecasts. Inset represents frequency-of-use diagram.

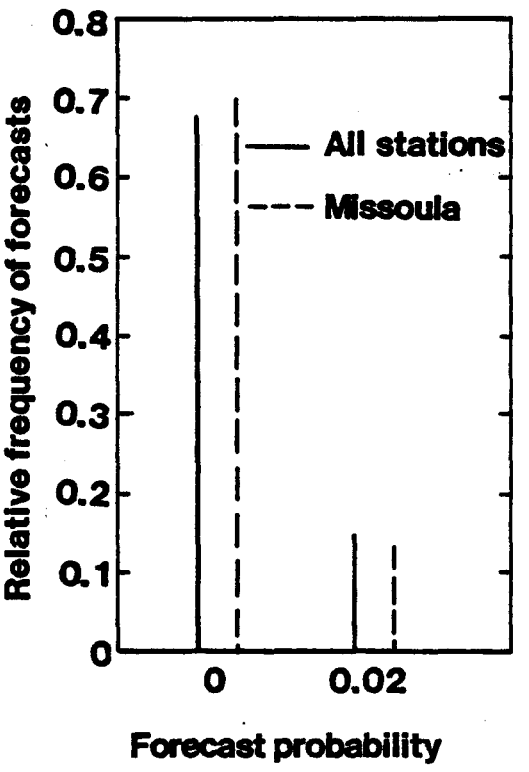

Figure 7. Frequency-of-use diagram for HLT20 forecasts.

probability values used also tends to decrease as the sample relative frequency decreases.

Measures of quality for the wind speed, wind direction, and relative humidity forecasts are presented in Tables 16, 17, and 18, respectively. The scores for the wind speed and relative humidity forecasts are based on the ranked probability score (RPS) and the wind direction scores are based on the Brier score (BS). In addition to the RPS or BS, these tables include the terms in the partition of RPS or BS - namely, the ranked probability or Brier score for sample climatology (RPS_{SC} or BS_{SC}), the reliability (REL), and the resolution (RES) - and the skill score (SS) based on the RPS or BS.

In the case of wind speed (Table 16), the skill scores are strongly negative for both all offices combined and Missoula alone. For example, for all offices combined, forecasts based solely on sample climatological probabilities are 85.9% more accurate than the experimental forecasts. The wind direction forecasts (Table 17) also exhibit negative skill, although the skill scores are not as strongly negative as they are for wind speed. For relative humidity (Table 18), the skill of the forecasts is also negative, with the skill score being somewhat lower (i.e., more negative) for all offices combined than for the Missoula forecasters alone. These results reflect the relatively large values of REL and the relatively small values of RES (see Section 4b).

Table 16. Ranked probability score (RPS), terms in partition of RPS (RPS_{SC}, REL, and RES), and skill score (SS) for wind speed forecasts.

Office	Sample size	RPS	RPS_{SC}	REL	RES	SS (%)
All	46	0.3769	0.2027	0.3552	0.1810	-85.9
Missoula	34	0.3673	0.2284	0.3526	0.2137	-60.8

Table 17. Brier score (BS), terms in partition of BS (BS_{SC}, REL, and RES), and skill score (SS) for wind direction forecasts.

Office	Sample size	BS	BS_{SC}	REL	RES	SS (%)
All	46	0.2124	0.1928	0.0500	0.0304	-10.2
Missoula	30	0.2347	0.2222	0.0658	0.0533	-5.6

Table 18. Ranked probability score terms in partition of RPS (RPS_{SC}, REL, and RES), and skill score (SS) for relative humidity forecasts.

Office	Sample size	RPS	RPS_{SC}	REL	RES	SS (%)
All	53	0.3818	0.3154	0.2529	0.1865	-21.0
Missoula	37	0.3656	0.3243	0.2890	0.2477	-12.7

6. Summary and Discussion

In this section we summarize the primary results of the probabilistic spot fire-weather forecasting experiment and discuss their implications. In addition, we consider some possible explanations for the relatively poor quality of the experimental forecasts. This discussion leads to some recommendations for future work in this area, and these recommendations are presented in Section 7.

It is difficult to judge the overall quality of the credible interval temperature forecasts since we do not have a standard of comparison such as climatology available. However, the experimental T50 forecasts were quite reliable, with only a small negative bias. Moreover, T50 corresponded reasonably well to the observed temperature (T). For example, the MAE and RMSE values comparing T50 and T were fairly small, ranging between 4 and 6°F. However, it is important to recall that the spot forecasts usually involved very short lead times (1/2 to 1 hour) and that the forecasters generally had access to a recent temperature observation from the burn site when formulating the forecast. These factors undoubtedly contributed significantly to the success of the T50 forecasts. The central credible interval forecasts were quite reliable overall, with 52% of the observed temperatures for all offices falling between T25 and T75. These intervals were slightly asymmetric, with the lower half of the interval (from T25 to T50) being somewhat narrower on the average than the upper half of the interval (from T50 to T75).

The observed relative frequencies of the wind speed and relative humidity events increased monotonically with increasing forecast probability, although all of these events were overforecast. With regard to the reliability of the wind speed and relative humidity forecasts, only the HLT40 forecasts appeared to be very reliable. In addition, a probability of zero was the most

frequently used value for the WGT15 forecasts, an event that did not occur in the experimental period. The wind direction forecasts were quite reliable overall. Skill scores for the wind speed, relative humidity, and wind direction forecasts were all negative, with particularly low scores associated with the wind speed forecasts. The latter scores equaled -85.9% for all offices combined and -60.8% for the Missoula forecasts.

One possible explanation for the extremely poor quality of the probabilistic wind speed forecasts is the difficulty associated with obtaining adequate wind speed verification measurements. In this regard, we were warned by the forecasters at all four offices at the outset of the experiment that it would be very difficult to obtain representative wind speed observations. Specifically, wind speed observations are critically dependent on measurement location and length of averaging time. Hence, the wind speed forecasts frequently may have involved a different event than the event that was actually observed. It is also extremely difficult to forecast wind speed in locations with complex terrain.

On the other hand, it evidently was less difficult to forecast wind direction and to verify such forecasts. The skill scores associated with these forecasts were only marginally less than zero and the forecasts were quite reliable overall. In this regard, wind direction in mountainous terrain frequently depends more on topography than on other factors. Thus, if a forecaster was aware that the wind was upslope at the time that the forecast was requested, it might be reasonable for him to assume that it would continue to be upslope at the verification time as well or to adjust the direction according to topographical or meteorological factors. The forecasters also expressed less uncertainty in the wind direction forecasts than in forecasts

of the other events, as indicated by the large probability values that were used in these forecasts.

Another factor that may have contributed to the overall poor quality of the experimental forecasts is the forecasters' relative lack of experience in formulating probabilistic forecasts. As noted in Section 2, this experience consisted only of the formulation of a small number of probability of precipitation forecasts. In addition, the forecasters were not provided with feedback regarding their performance during the experiment. Such feedback might have enabled the forecasters to make more reliable and skillful forecasts during the latter part of the experimental period (e.g., see Murphy and Daan, 1984).

7. Conclusions and Recommendations

As discussed in Section 2, resource managers such as those charged with planning and carrying out prescribed burns have a need for information regarding the uncertainty in weather forecasts. The weather is a critical factor in determining the outcome of a prescribed burn and must be considered at all stages of planning and monitoring the burn. It is particularly important to have information available regarding the weather at the planned ignition time since the weather conditions at that time have a significant impact on the success of the burn. Moreover, in planning a prescribed burn, limits are placed on the acceptable ranges of values of the critical weather variables. Probabilistic spot fire-weather forecasts would provide essential information regarding the likelihood that the values of the variables will fall within the critical ranges of the prescribed burn plan (this information is not contained in current categorical spot fire-weather forecasts). Hence, probabilistic forecasts would give the burn manager additional information regarding the acceptability of the future weather conditions that should be quite useful in deciding whether or not to proceed with a planned burn.

In summary, it appears to be desirable to provide burn managers with information regarding the uncertainty associated with spot fire-weather forecasts. For this reason, we believe that it would be useful to undertake further studies of the ability of fire-weather forecasters to quantify the uncertainty in their spot forecasts and, ultimately, to initiate an operational probabilistic spot fire-weather forecasting program. Despite the somewhat disappointing outcome of the forecasting experiment described in this report, we are quite confident that additional experiments would yield more positive results. However, it obviously would be desirable to take the experience gained in this study into account in planning future experiments,

so that the latter will be more successful and yield more useful results. The study reported here suffered from several operational difficulties (e.g., the collection of verification data) which could be remedied in future experiments. Moreover, the lack of probabilistic forecasting experience by the forecasters involved in the experiment could be alleviated by having the forecasters participate in a series of experiments and by providing them with feedback during the course of (or between) the experiments.

In particular, we have the following recommendations regarding future experiments related to the formulation of probabilistic spot fire-weather forecasts [the reader may also want to refer to the recommendations regarding additional studies of probabilistic fire-danger weather forecasts made by Brown and Murphy (1985)]:

(a) The experiments should be conducted in several locations involving different climatological and meteorological regimes.

(b) Verification data are of critical importance to the success of a forecasting experiment. The experiment described in this report suffered greatly from the fact that only small number of matched forecasts and observations were obtained. Therefore, based on this experience, it is of particular importance to ensure that verifying measurements are made for every experimental forecast and that these measurements are made according to a standard method for each forecast. In order to be certain that the appropriate observations are made for each forecast, it may be necessary to have an individual in each National Forest dedicated to obtaining the relevant measurements throughout the experimental period.

(c) The forecasters should be provided with feedback regarding their individual and collective performance at appropriate times within

and between experiments. Such feedback can be invaluable in helping the forecasters to improve their probabilistic forecasts.

(d) The experiments should be carefully evaluated and the usefulness of the forecasts in the context of decision-making for prescribed burns should be investigated.

If these experiments yield positive results, then careful consideration should be given to an operational program of probabilistic spot fire-weather forecasts. Several steps would be involved in designing and implementing such a program, and Murphy and Winkler (1979) have briefly described some of these steps in the context of a proposed operational program of probabilistic temperature forecasting.

Acknowledgments

The experiment described in this report would not have been possible without the interest and participation of the forecasters at the Boise, Medford, Missoula, and Wenatchee National Weather Service offices. We greatly appreciate the forecasters' help in designing the experiment (in particular, in selecting the appropriate events to include in the experimental forecasts) and their willingness to add the experimental probability forecasts to an already demanding set of duties. In addition, we would like to thank the Forest Service personnel who recorded the verifying observations that were an essential part of the experiment. We would also like to thank Mr. William Parker for his assistance in initiating contact with the NWS offices and Mr. David Radloff for his help in contacting individuals in the National Forests to aid in the collection of verification data.

Dr. Peter Roussopoulos provided helpful comments on a draft version of this report.

This research was supported by the Forest Service, U.S. Department of Agriculture, under Research Contract No. 28-K3-302 with the Rocky Mountain Forest and Range Experiment Station.

References

Brown, B.G., and A.H. Murphy, 1984: Current procedures and practices in spot fire-weather forecasting. Corvallis, Oregon State University, Department of Atmospheric Sciences, SCIL Report 84-1 (Report #1 to USDA - Forest Service), 43 pp.

Brown, B.G., and A.H. Murphy, 1985: Quantification of uncertainty in operational and experimental fire-danger forecasts. Corvallis, Oregon State University, Department of Atmospheric Sciences, SCIL Report 85-1 (Report #3 to USDA - Forest Service), 220 pp.

Epstein, E.S., 1969: A scoring system for probability forecasts of ranked categories. Journal of Applied Meteorology, 8, 985-987.

Murphy, A.H., 1971: A note on the ranked probability score. Journal of Applied Meteorology, 10, 155-156.

Murphy, A.H., and H. Daan, 1984: Impacts of feedback and experience on the quality of subjective probability forecasts: comparison of results from the first and second years of the Zierikzee experiment. Monthly Weather Review, 112, 413-423.

Murphy, A.H., and H. Daan, 1985: Forecast evaluation. Probability, Statistics, and Decision Making in the Atmospheric Sciences (A.H. Murphy and R.W. Katz, Editors). Boulder, Colorado, Westview Press, pp. 379-437.

Murphy, A.H., and R.L. Winkler, 1979: Probabilistic temperature forecasts: the case for an operational program. Bulletin of the American Meteorological Society, 60, 12-19.

Radloff, D.L., and R.F. Yancik, 1983: Decision analysis of prescribed burning. Preprints, Seventh Conference on Fire and Forest Meteorology (Fort Collins). Boston, Massachusetts, American Meteorological Society, pp. 85-89.

Winkler, R.L., and A.H. Murphy, 1979: The use of probabilities in forecasts of maximum and minimum temperatures. Meteorological Magazine, 108, 317-329.

Appendix

Forecast Aids and Forms Provided to Forecasters Participating in Probabilistic Spot Fire-Weather Forecasting Experiment

PROBABILISTIC SPOT FIRE WEATHER FORECASTING EXPERIMENT

Aids in Assessing Probabilities and Checking their Consistency

This document contains some suggestions of assessment methods and consistency checks that may be helpful to you for preparing the probability forecasts for this experiment. The methods are described in detail for forecasting wind speed, with briefer discussions of the wind direction and humidity forecasts. The discussion of temperature forecasts is somewhat more detailed since the nature of the temperature probability forecasts is a bit different from the other types of forecasts. For all of the elements, the subjective probability forecasts should reflect your best judgment -- based on all available evidence and information -- regarding the likelihood of occurrence of the events in question.

A. Wind speed

The probability forecast of wind speed (WS) consists of probabilities associated with three events:

1. WS > 5
2. WS > 10
3. WS > 15

For convenience, we will denote these events by e_5, e_{10}, and e_{15}, respectively. Furthermore, we will denote your subjective probability forecasts for these events by P_5, P_{10}, and P_{15}, respectively (e.g., P_5 is your subjective probability assessment of the likelihood that the wind speed at valid time will be greater than 5 mph).

In quantifying your uncertainty concerning the likelihood of occurrence of

specific events, it may sometimes be useful to consider simple betting situations. Suppose you are assessing P_5 (i.e., $Pr\{WS > 5\}$). Consider the following specific bet, in which you must choose between A and B:

A: receive \$X for certain (an outright gift)

B: receive \$10 if WS > 5.

Suppose you are indifferent between A and B for some value of X (i.e., you do not prefer A to B or vice versa). This indifference implies that $10 \times P_5 = X$ or $P_5 = X/10$. Thus, if X = \$4 on some particular occasion, it implies that $P_5 = 0.4$ (for you) on that occasion. Similar betting situations can be used to help you assess P_{10} and P_{15}.

After you have assessed P_5, P_{10}, and P_{15}, it is desirable to perform certain checks to ensure the consistency of these probabilities and to make certain that they actually reflect your best judgments about the events. First, since the event e_{10} is contained in the event e_5, and the event e_{15} is contained in e_{10}, the following relationship among the probabilities obviously must hold:

$$P_5 \geq P_{10} \geq P_{15}.$$

If these inequalities are violated, you should reassess P_5, P_{10}, and P_{15}.

A second check involves looking at events that are complementary to e_5, e_{10}, and e_{15}. When you report that $Pr\{WS > 5\} = P_5$, you are implying that $Pr\{WS \leq 5\} = 1 - P_5$. Is this latter probability consistent with your best judgments? If not, then you should reassess P_5. Similar consistency checks can be performed for P_{10} and P_{15}.

A third consistency check can be performed based on non-overlapping events. Your assessments P_5, P_{10}, and P_{15} imply that

$$\Pr\{WS \leq 5\} = 1 - P_5$$

$$\Pr\{5 < WS \leq 10\} = P_5 - P_{10}$$

$$\Pr\{10 < WS \leq 15\} = P_{10} - P_{15}$$

$$\Pr\{WS > 15\} = P_{15}.$$

Are these probabilities of non-overlapping events consistent with your judgments? If not, you should reassess P_5, P_{10}, and P_{15}.

In summary, you can assess your subjective probabilities (P_5, P_{10}, and P_{15}) of the three wind speed events (e_5, e_{10}, and e_{15}) using simple betting situations. Once you have assessed these probabilities, it is desirable to perform the following three checks of the consistency of the probabilities.

1. Is $P_5 \geq P_{10} \geq P_{15}$?
2. Do the values $1 - P_5$, $1 - P_{10}$, and $1 - P_{15}$ properly represent your judgments regarding the probabilities associated with the events $WS \leq 5$, $WS \leq 10$, and $WS \leq 15$?
3. Do the values P_5, $P_5 - P_{10}$, $P_{10} - P_{15}$, and P_{15} properly represent your judgments regarding the probabilities associated with the non-overlapping events $WS \leq 5$, $5 < WS \leq 10$, $10 < WS \leq 15$, and $WS > 15$, respectively?

If your response to any of these checks is "no", you should reassess P_5, P_{10}, and P_{15}.

B. Wind direction

Only one event is of concern for the probability forecast of wind direction (WD):

$$\overline{WD} - 45° < WD < \overline{WD} + 45°,$$

where $\overline{WD}$ is your best estimate of the wind direction at valid time to the nearest of eight cardinal compass points. We will denote this event by e and your subjective probability forecast of the event by P. Once you have estimated $\overline{WD}$, P can be assessed using a betting situation similar to that described for the wind speed events. To assess P, find the value of X for which you are indifferent between the two bets A and B:

A: receive \$X for certain

B: Receive \$10 if $\overline{WD} - 45° < WD < \overline{WD} + 45°$.

Then your subjective asssessment of P is X/10.

Because the wind direction forecast involves only one event, only the second of the three consistency checks described for the wind speed forecasts is applicable. That check is the check for complementary events. When you report that $Pr\{\overline{WD} - 45° < WD < \overline{WD} + 45°\} = P$, you are implying that the probability that WD is <u>not</u> in that region is equal to 1 - P. If this latter probability is not consistent with your best judgments you should reassess P.

C. <u>Humidity</u>

The probability forecast of humidity (H) consists of probabilities associated with three events:

1. $H < 40$
2. $H < 30$
3. $H < 20$

We will denote these events by e_{40}, e_{30}, and e_{20}, respectively, and your probability forecasts of the events by P_{40}, P_{30}, and P_{20}, respectively. These probabilities may be assessed in a manner analogous to the betting procedure described for probability forecasts of wind speed. For example, to assess P_{30}

find the value of X for which you are indifferent between the two bets A and B:

A: receive \$X for certain

B: receive \$10 if $H < 30$.

Then P_{30} on that occasion equals X/10.

Three checks of the consistency of your humidity probability forecasts are desirable. These checks correspond to the checks described for the wind speed forecasts and are as follows:

1. Is $P_{40} > P_{30} > P_{20}$?
2. Does $Pr\{H > 40\} = 1 - P_{40}$ represent your best judgments about the event $H > 40$?

   Does $Pr\{H > 30\} = 1 - P_{30}$ represent your best judgments about the event $H > 30$?

   Does $Pr\{H > 20\} = 1 - P_{20}$ represent your best judgments about the event $H > 20$?
3. Do the non-overlapping event probabilities

$$Pr\{H < 20\} = P_{20}$$
$$Pr\{20 < H < 30\} = P_{30} - P_{20}$$
$$Pr\{30 < H < 40\} = P_{40} - P_{30}$$
$$Pr\{H > 40\} = 1 - P_{40}$$

   represent your best judgments about these events?

If you answer "no" to any of these questions you should reassess the probabilities P_{40}, P_{30}, and P_{20}.

D. Temperature

In forecasting temperature at valid time, you undoubtedly are somewhat uncertain about what the actual temperature will be. It is possible to give a

point forecast (i.e., a single value) that represents your "best estimate" about the temperature, but the point forecast alone does not completely represent your uncertainty. A convenient way to convey this uncertainty is through the use of interval forecasts (i.e., intervals of values, as opposed to the single values used as point forecasts). Specifying an interval and the probability that the temperature will be within the interval conveys a considerable amount of information about your uncertainty. In some situations, you may feel that the odds are even that the temperature will be in a particular five-degree interval; on other days, you may be much more uncertain, so you feel that the odds are even that the temperature will be in a particular ten-degree interval. In this experiment you will be asked to determine an interval such that the probability is 50% that the temperature at valid time will be in the interval. An interval is assumed to include its end points; for example, the interval 72-76°F is a five-degree interval (it includes 72, 73, 74, 75, and 76). Note that in determining your interval forecasts, you will be working with intervals that are of fixed probability (50%), and you will have to determine the end points of the intervals; hence, the intervals are of variable width (the width depending on how uncertain you are on a given occasion).

The first step in determining the interval forecasts is to determine a median, which will be used as a mid-point for the variable width interval. A median is a value that you feel is equally likely to be exceeded or not exceeded. For example, if you feel that it is equally likely that the temperature at valid time will be above 74 or below 74, then 74 is your median. The following dialogue should illustrate how you might arrive at a median.

Experimenter: What is your best intuitive estimate of the temperature?

Forecaster: About 90 degrees.

Experimenter: My first step will be an attempt to sharpen up that initial estimate. If we were both to wager the same amount of money, would you rather bet that the temperature will be above 90 degrees or below?

Forecaster: Above 90 degrees.

Experimenter: Would you rather bet that it will be above 94 degrees or below?

Forecaster: Below.

Experimenter: Above or below 91 degrees?

Forecaster: Hmmm .... probably above.

Experimenter: Above or below 92 degrees?

Forecaster: It doesn't make much difference there.

Experimenter: Above or below 93 degrees?

Forecaster: Below.

Experimenter: Fine. Then we will select 92 degrees as your indifference judgment. You think that it is just as likely that the temperature at valid time will be above 92 degrees as that it will be below 92 degrees. Is that right?

Forecaster: That seems right.

Experimenter: In a sense, 92 degrees, which is a median, is your best estimate of the temperature at valid time -- it can be viewed as a point forecast.

The next step is to determine your 25th percentile (the median is sometimes called the 50th percentile). The 25th percentile is the value that divides the interval <u>below the median</u> into two equally likely sub-intervals. Note that the median divided the entire set of possible values into two equally likely intervals, so the procedure for determining the 25th percentile is very similar to the procedure for determining the median. For example, suppose that your median for the temperature at valid time is 74. Then if you feel that it is equally likely that the temperature will be below 71 or between 71 and 74, then 71 is your 25th percentile. The following continuation of the dialogue presented above illustrates the determination of a 25th percentile.

Experimenter: In a sense, 92 degrees, which is a "median", is your best estimate of the temperature at valid time. The next series of questions that I'll ask is designed to explore just how certain you are that the temperature will be near 92 degrees. First, assume that all bets are off in case the temperature is greater than 92 degrees. Do you think that it is more likely that the temperature at valid time will fall below 80 degrees or between 80 and 92 degrees.

Forecaster: It is more likely to be between 80 and 92 degrees.

Experimenter: Below 85 degrees, or between 85 and 92 degrees?

Forecaster: That's pretty difficult. Probably below 85 degrees.

Experimenter: Below 84 degrees or between 84 and 92 degrees?

Forecaster: That's about it. I can't choose between the two intervals.

Experimenter: Fine -- then we will accept 84 degrees as your 25th percentile.

The next step is to determine your 75th percentile, the value that divides the interval <u>above the median</u> into two equally likely subintervals. As you might suspect, the procedure for determining the 75th percentile is like the procedure for determining the 25th percentile. Let's go back to the dialogue.

Experimenter: Now let's move on to the upper range, the range above the median. Assuming that all bets are off if the temperture at valid time is below 92 degrees, do you think that it is more likely to be between 92 and 100 or above 100?

Forecaster: Definitely between 92 and 100.

Experimenter: Between 92 and 95 or above 95?

Forecaster: Still between 92 and 95.

Experimenter: Between 92 and 94 or above 94?

Forecaster: Now I am indifferent.

Experimenter: In that case we will take 94 as your 75th percentile.

The median, the 25th percentile, and the 75th percentile have been determined, in that order. These values can be used to determine an interval forecast. The probability is 50% that the temperature at valid time will be

between the 25th percentile and the 75th percentile. It is useful to reconsider the values that have been determined to make sure that they coincide with your best judgments. To illustrate this, we return to the dialogue one more time.

Experimenter: Now let's carefully consider the values that you have estimated. First, consider the intervals A, B, C, and D, where A is below 84 degrees, B is between 84 and 92, and C is between 92 and 94, and D is above 94. Assume that there is a four-way bet this time and you can pick only one of the intervals. Which one would you prefer?

Forecaster: Hmmm ... clearly not B or C. I guess I like A the best, but D looks pretty good, too.

Experimenter: People occasionally squeeze the outside boundaries in too closely when making judgments like this for the first time.

Forecaster: I must have done that because now I clearly like the outside two intervals better than the middle ones.

Experimenter: Then move the outer boundaries out one degree each so that the boundaries are at 83 degrees, 92 degrees, and 95 degrees. Now which interval would you prefer to bet on?

Forecaster: These estimates are better now. Any one of the intervals looks just as good as any other one to me. Also, I think that the temperture is just as likely to fall inside the interval between 83 and 95 degrees as it is to fall outside that interval.

Experimenter: Good. Then your interval estimate with probability 50% is from 83 to 95. It is interesting that the boundaries are spread out asymmetrically around 92 degrees. The lower bound of 83 degrees has been pushed much farther away than the upper boundary of 95 degrees.

Forecaster: I was thinking about that when making my estimates. A weak cold front is moving in from the northwest. It may reach here early tomorrow morning, but it may take until tomorrow night. If it gets here before morning, then it won't get very warm tomorrow, but, if the front is delayed, then the temperature should be around 92 degrees.

Experimenter: Then that explains why the upper boundary is so much closer to 92 degrees. There is little chance for any change in conditions to produce much of an increase above your median of 92.

Forecaster: That's right. Looked at that way, these intervals display a lot of what I know about tomorrow's temperature. They don't indicate why the temperature could drop but they certainly show that it can. I wouldn't expect to always have such asymmetric intervals when compared with the median, but it sure seems reasonable in this particular situation.

For convenience, here is a summary of the procedure:

1. Determine your median ($T_{.50}$).

2. Determine your 25th percentile ($T_{.25}$).

3. Determine your 75th percentile ($T_{.75}$).

4. Look at the resulting intervals to make sure that they agree with your judgments, making any changes you deem necessary.

BOISE

PROBABILISTIC SPOT FIRE-WEATHER FORECASTING EXPERIMENT

Forecast Form

Date (month/day) _____

Forecast time (time at which forecast is made/issued) _____

Name of fire/project _____

Fire/project number (if applicable) _____

Forest/ranger district _____

Valid time (time at which forecast is valid and at which ignition is expected) _____

Forecaster identification _____

Probability Forecasts

In items (1), (2), and (3), give probabilities to nearest 10%, except that values of 2%, 5% 95%, and 98% also are permitted. In item (4), give temperatures to nearest °F. See instruction sheet on reverse side for additional information.

(1) Wind speed (WS) (eye-level wind speed in mph)

Pr(WS > 5) _____ Pr(WS > 10) _____ Pr(WS > 15) _____

(2) Wind direction (WD) (eye-level wind direction to nearest of 8 cardinal compass points)

$\overline{WD}$ _____

Pr($\overline{WD}$ - 45° < WD < $\overline{WD}$ + 45°) _____

($\overline{WD}$ is the forecaster's best estimate of the wind direction at valid time)

(3) Humidity (H) (relative humidity in percent)

Pr(H < 40) _____ Pr(H < 30) _____ Pr(H < 20) _____

(4) Temperature (T) (air temperature in °F)

$T_{.50}$ _____ $T_{.25}$ _____ $T_{.75}$ _____

(see instruction sheet for explanations of $T_{.50}$, $T_{.25}$, and $T_{.75}$)

PROBABILISTIC SPOT FIRE-WEATHER FORECASTING EXPERIMENT

Instructions for Forecasters

Date: Date on which forecast is made/issued (month/day).

Forecast time: Time at which forecast is made/issued (to nearest hour and minute local time).

Name of fire/project: Name given to fire/project by individual requesting forecast.

Fire/project number: Number assigned to fire/project (may not be applicable in some offices).

Forest/ranger district: National Forest and ranger district for which forecast is requested.

Valid time: Time at which forecast is valid and at which ignition is expected (to nearest hour in local time).

Forecaster identification: Identification number assigned to forecaster for this experiment.

Probability Forecasts

(1) Wind speed: (WS) Provide your best judgments in terms of numerical probabilities that the observed eye-level wind speed at valid time (expected ignition time) will be (a) greater than 5 mph, Pr(WS > 5); (b) greater than 10 mph, Pr(WS > 10); and (c) greater than 15 mph, Pr(WS > 15). Note that, according to the definitions of these events, Pr(WS > 5) > Pr(WS > 10) > Pr(WS > 15). Values of 0%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, and 100% may be used for these probabilities.

(2) Wind direction (WD): First determine and record your best estimate of the eye-level wind direction (to the nearest of 8 cardinal compass points) at valid time ($\overline{WD}$). Then provide your best judgment in terms of numerical probabilities that the observed eye-level wind direction at valid time will be within $\mp$ 45° of $\overline{WD}$ - that is, Pr($\overline{WD}$ - 45° < WD < $\overline{WD}$ + 45°). For example, if your best estimate is that the winds will be from the southwest (225°), then the forecast provided should be the probability that the wind direction will be between south (180°) and west (270°). Values of 0%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, and 100% may be used for this probability.

(3) Humidity (H): Provide your best judgments in terms of numerical probabilities that the observed relative humidity at valid time will be (a) less than 40%, Pr(H < 40); (b) less than 30%, Pr(H < 30%); and (c) less than 20%, Pr(H < 20%). Note that, according to the definitions of these events, Pr(H < 40%) > Pr(H < 30%) > Pr(H < 20%). Values of 0%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, and 100% may be used for these probabilities.

(4) Temperature (T): Provide your best judgments of the median ($T_{.50}$), lower quartile ($T_{.25}$), and upper quartile ($T_{.75}$) of your subjective probability distribution for the observed air temperature at valid time. If the latter is denoted by T_0, then $T_{.50}$ is the temperature for which the chance of T_0 being below $T_{.50}$ is equal to the chance of T_0 being above $T_{.50}$ (with each probability equal to 50%). The interval between $T_{.25}$ and $T_{.75}$ is the range of temperature values in which you believe there is a 50% chance of T_0 falling, with the probability of T_0 being less than $T_{.25}$ equal to 25% and the probability of T_0 being greater than $T_{.75}$ also equal to 25%.

Please refer to the forecasting aids provided for this experiment to obtain additional information for assessing and checking these probabilities and temperatures.